Fast MC simulation of top events

S.Chekanov (ANL)

April, 2013

+ discussion with A. Schwartzman. C.Brock, T.LeCompte, J.Proudfoot, A.Kotwal, S.Padhi,

Why?

- **During long range planning exercise at Snowmass, we need to make physics cases for high luminosity runs**
  - Need for a simulation platform which goes beyond the scopes of each separate experiments
- **The simulation platform cannot be based on tools and software of separate experiments**
  - Software developed by collaborations cannot be used outside collaborations
  - Should easily be accessible by everyone (including theorists)
  - The usage of grid should be minimal (no “Snowmass VO”!)
- **Simulation toolkit should be easy to run, and it should fast!**
  - Rules out Geant-based full simulations
  - A combination of publicly available MC generators with fast simulations is optimal

Is fast simulations realistic?

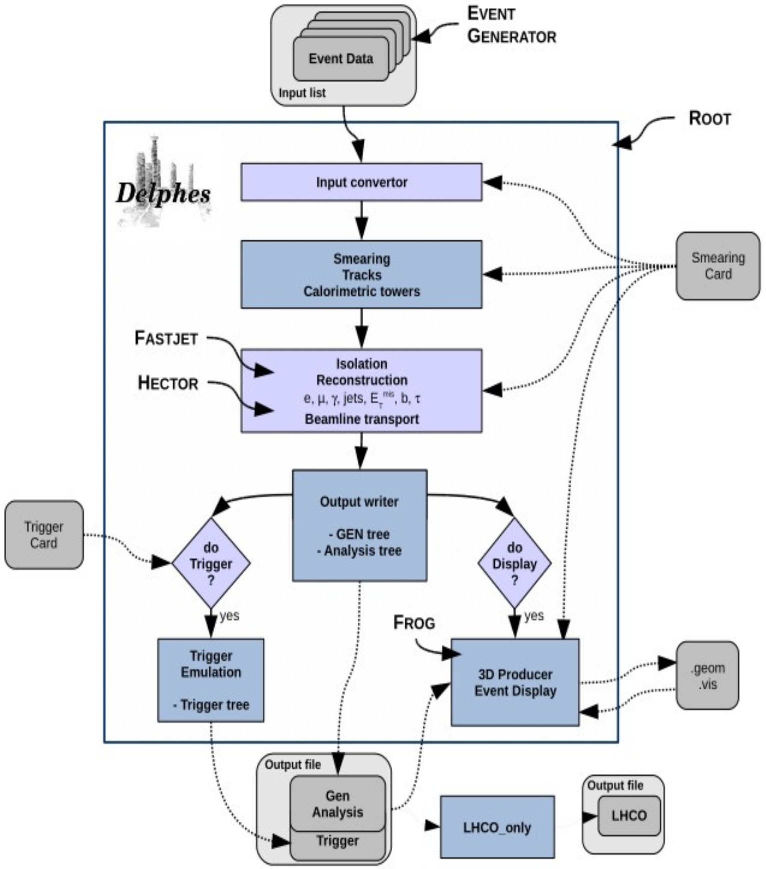
- **Is fast simulations realistic?**
  - It cannot be fully realistic, but it should catch the main features of each detector
  - A validation / comparison with the full simulation is essential

- **Is the full detector realistic?**
  - when it comes to physics using high-lumi runs, large contributor is not "hard-wired" detector instrumentation, but methods dealing with pileup events (see later)
  - Full detector simulation can only be realistic with full understanding of pileup removal methods, calibrations, removal of dead cells, i.e. in combination with a complete analysis infrastructure
  - Modeling of soft events at ~14 TeV is not obvious and requires extrapolation using 8-TeV UE models and certain assumptions
  - To make a comprehensive survey of physics analysis for different pileup scenarios using full simulation beyond reach of each separate experiments

- **Fast simulation will come in handy:**
  - Build a comprehensive survey of various physics channels using minimum computing and intellectual resources
  - Helps to study pileup effects & develop pileup subtraction techniques

Delphes fast simulations
(https://cp3.irmp.ucl.ac.be/projects/delphes)

- **A lot of recent studies went into "Delphes" MC fast simulation**
  - Line by line comparison of Delphes2 and Delphes3 indicates that the implementations of ATLAS and CMS in these cards are identical (small problems were reported to the Delphes team)
  - Smearing prescriptions exist using Delphes2 ATLAS cards for 140 pileups (Tom) → can directly be used in Delphes3
- **Delphes3 looks more promising than Delphes2 since Delphes3 is well supported and can mix signal and pileup events**
  - Many improvements for Delphes3 were motivated by this study



Snowmass detector

- **Decided to build an idealized Delphes input card based on CMS and ATLAS detectors**
  - Difference between CMS and ATLAS detector descriptions may have a second-order effect on final physics outcome for high-luminosity runs (need to check this!).
  - Different techniques to reduce pileup effects on missing ET and jets become essential contributor to physics cases
  - If we fail to build physics case for certain processes using “best” detector features of ATLAS and CMS, we likely will fail when using full simulations or real data (the opposite may not be true!)
  - It should also be recognized that difference in analysis methods may have larger effect on the final physics outcome compared to differences in detector instrumentation
  - We are less interested in how well CMS or ATLAS detector look like under different pileup conditions. We are interested in how physics may look for high-luminosity runs using certain simplification and assumptions
- **A geometry card was developed based on best performance of each detector (ATLAS and CMS)**

MC samples with pileup events

- **Build a framework to generate signal events and validate them using full detector simulations (if possible)**
- **Generate signal events with several pileup scenario**
- **We will have both “parametric” and “realistic” Delphes samples**
- **Unlike “parametric” simulation of pileup, signal+pileup fast simulation can be quite realistic**
  - have extra fakes which are missing when smearing jets or reconstructed leptons
  - certain effects cannot be simulated using smearing functions (event shapes, jet substructure, jet masses etc)
  - useful when comparisons with full.sim. are missing
- **Less concentrate on possible differences in underlying kinematics (NLO vs LO). Main questions:**
  - how well can we reconstruct top quark for different pileup scenarios?
  - what uncertainties are expected?
  - what physics processes are less / more affected by pileups?

MC samples for download

Below are ttbar and background MC samples generated for pp collisions at 14 TeV

1. **Delphes2.03 for PYTHIA8 and HERWIG++ for high-pT ttbar and high-pT QCD jets (without pileup)**
2. **Delphes3.0 beta for PYTHIA8. ttbar without pileup**

14 TeV pp events, Delphes3.05 and ATLAS-like geometry card with antiKT R=0.5 ("pflow" as input)

1. **HERWIG++ for low pT ttbar signal and signal+pileup (140 soft events)**
2. **HERWIG++ for high pT ttbar (>650 GeV) signal and signal+pileup (140 soft events)**
3. **HERWIG++ pileup events for the signal events used in (1) and (2)**
4. **HERWIG++ for low pT ttbar signal and signal+pileup (50 soft events)**
5. **HERWIG++ for high pT ttbar (>650 GeV) signal and signal+pileup (50 soft events)**
6. **PYTHIA8 for Zprime(3000) to ttbar with mu=0,50,140 pileups** New
7. **PYTHIA8 for single top (s- and t- channels) with mu=0,50,140 pileups** New

+ only pileup samples (50 or 140 soft events per pp collision)

Each sample has ~20,000 signal events
Each event is mixed with either 50 and 140 events soft events (HERWIG++)
(done using ANL HEP Tier3)

https://atlaswww.hep.anl.gov/snowmass13

Truth record: HERWIG++ 2.6.3 for $t\bar{t}$ (pp, 14 TeV)

- HERWIG++ default tuning. Low pT $t\bar{t}$. No pile-up

  - set /Herwig/Cuts/JetKtCut:MinKT **10.0*GeV**
  - ## This should be <= 2 * JetKtCut:MinKT unless you *want* a mhat cut. Default is 20 GeV.
  - set /Herwig/Cuts/QCDCuts:MHatMin **20.0*GeV**
  - # Colour reconnection settings
  - set /Herwig/Hadronization/ColourReconnector:**ColourReconnection Yes**
  - set /Herwig/Hadronization/ColourReconnector:ReconnectionProbability 0.6165547
  - # Colour Disrupt settings
  - set /Herwig/Partons/RemnantDecayer:colourDisrupt 0.3493643
  - # inverse hadron radius
  - set /Herwig/UnderlyingEvent/MPIHandler:InvRadius 0.81

- 64,000 ttbar events

Truth record: HERWIG++ 2.6.3 for $t\bar{t}$ (pp, 14 TeV)

- HERWIG++ default tuning. High-PT $t\bar{t}$. No pile-up
  - set /Herwig/Cuts/JetKtCut:MinKT **650.0*GeV**
  - ## This should be <= 2 * JetKtCut:MinKT unless you *want* a mhat cut. Default is 20 GeV.
  - set /Herwig/Cuts/QCDCuts:MHatMin 1200.0*GeV
  - # Colour reconnection settings
  - set /Herwig/Hadronization/ColourReconnector:**ColourReconnection Yes**
  - set /Herwig/Hadronization/ColourReconnector:ReconnectionProbability 0.6165547
  - # Colour Disrupt settings
  - set /Herwig/Partons/RemnantDecayer:colourDisrupt 0.3493643
  - # inverse hadron radius
  - set /Herwig/UnderlyingEvent/MPIHandler:InvRadius 0.81
- 64,000 ttbar events

Truth record: HERWIG++ 2.6.3 for $t\bar{t}$ (pp, 14 TeV)

- HERWIG++ default tuning. High-PT $t\bar{t}$. No pile-up
  - set /Herwig/Cuts/JetKtCut:MinKT **650.0*GeV**
  - ## This should be <= 2 * JetKtCut:MinKT unless you *want* a mhat cut. Default is 20 GeV.
  - set /Herwig/Cuts/QCDCuts:MHatMin 1200.0*GeV
  - # Colour reconnection settings
  - set /Herwig/Hadronization/ColourReconnector:**ColourReconnection Yes**
  - set /Herwig/Hadronization/ColourReconnector:ReconnectionProbability 0.6165547
  - # Colour Disrupt settings
  - set /Herwig/Partons/RemnantDecayer:colourDisrupt 0.3493643
  - # inverse hadron radius
  - set /Herwig/UnderlyingEvent/MPIHandler:InvRadius 0.81
- 64,000 ttbar events

How to download MC samples

ROOT files are available from the ANL server (~10 Gb/s)
https://atlaswww.hep.anl.gov/snowmass13

Use the "download.py" script to copy any number of ROOT files
Example: download 5 files with PYTHIA $t\bar{t}$ (pT>650 GeV):

```
python download.py 5 pythia8/ttbar650pt pythia8_ttbar_pt650
```

Nr of files to download

Directory

Generic name

(can stop it as [Ctrl]-[C] and restart it an any time)

Do not try to download all files (~80). Try first a few files

No grid registration

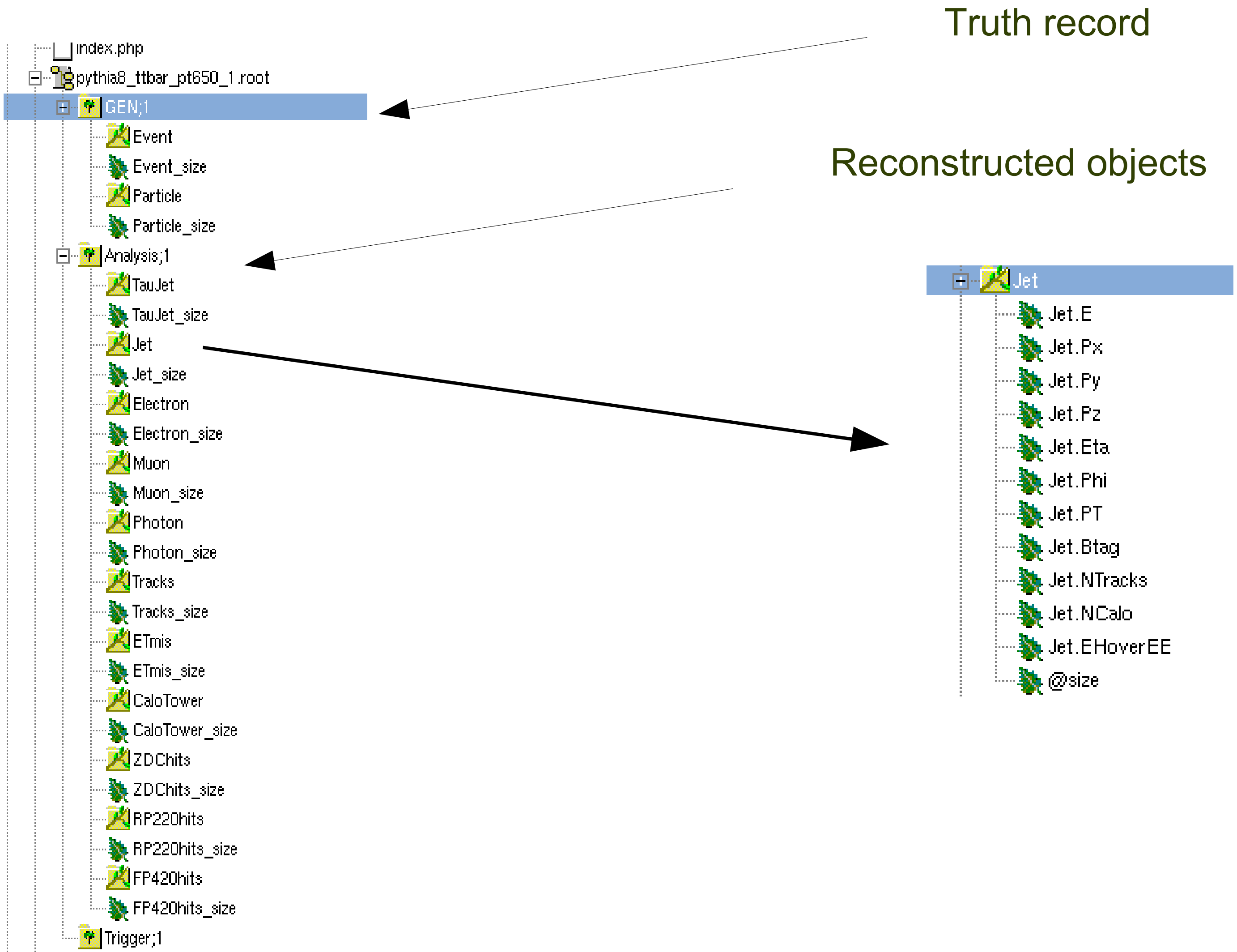
How to analyze

Get a few files and open them in TBrowser to see what is inside

>> root
>>TBrowser a

- **A more complicated C++ program which reads all ROOT files from a given directory is posted on the web**
- **Note:**
  - The program tightly integrated with the Delphes libraries. Delphes should be installed
  - Also Delphes 2.03 and 3.05 are quite different and need to be compiled separately
  - Look at the examples in the directory examples/*.C

Look at the structure:



Help

https://atlaswww.hep.anl.gov/snowmass13

Help for analyzers

- **How to download** link shows an automatic way of copying the ROOT files from this server
- **How to analyze** link explains how to analyze Delphes3 ROOT format
- **What slims are used** link explains what was done to reduce Delphes output
- **Accessing truth particles from signal events** link explains how to access truth particles ignoring pileup
- **How to generate MC samples** link explains how to generate fast simulation events with pileups
- **How to access objects** link explains how to access jets, leptons, missing ET
- **How to correct jets for pileup** link explains how to apply offset correction for pileup

Help for Delphes developers

You can add a new deatures to Delphes using this test HEPMC record. Download a HEPMC file with ttbar (>650 GeV, 14 TeV) for Delphes. Get it from here LHC_signal_ttbar650.hepmc.gz. Test Delphes as:

The page provides a tar file with the complete simulation framework. It includes:

- a bash script to compile (+ environment setting script)
- HERWIG++
- Delphes 3.05 (Fastjet is included)
- HEPMC
- PILEMC (from HepForge)
- HEPMC slim program (from Andy Buckley with small modification from me)

Compilation takes ~ 40 min (tested on SL5, PE710). You need ROOT!

"Slimmed truth"

- **Output MC files are large since have pileup particles!**
- **Steps to generate the files:**
  - Generate 200 signal events and save them in HEPMC ASCII files
  - Generate 30,000 soft events and save in HEPMC records (~1.2 GB)
  - Merge each signal event with 140 soft events using 30,000 soft as event pool
  - Process 200 mixed events with pileup using Delphes
  - Typical size is 800 MB for each 200 signal events
- **In the current implementation we "slim" MC record before Delphes3 step:**
  - We keep only stable particles and most essential particles needed for analysis (t,W,b,Z')
  - Remove tower jets
  - Mark pileup stable particles with the status code -1
- **The output ROOT file is 160 MB for 200 signal events with 140 pileups**

Many uses

- **MC samples can be used to validate fast simulations**
- **Construct additional resolution smearing for "parametric" simulations for different pileup scenarios (without mixing with pileups). In future.**
- **Samples can be used to study:**
  - Reconstruction of $t\bar{t}$ (for any decay channel)
  - Reconstruction of single top (t- and s- channels)
  - W mass
  - Study muons, electrons isolation
  - Study b-tagging
  - Resolution studies for jets and missing ET
  - Reconstruct jet masses (boosted top)
  - Study dijet masses
  - Z' reconstruction (M=3000 GeV)

- **Develop pileup subtraction techniques and look at concrete physics cases!**

First results

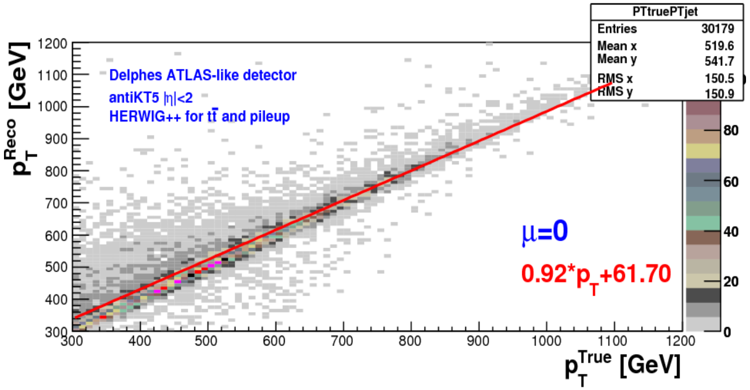
Note:

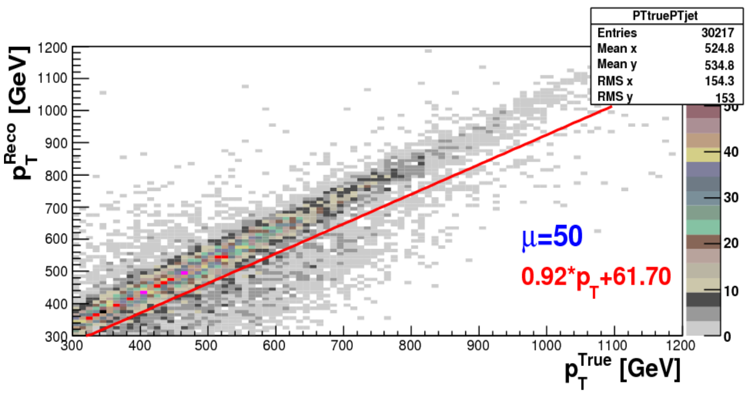
- **we do not apply “pileup” removal methods (they will be included with “Snowmass” samples)**

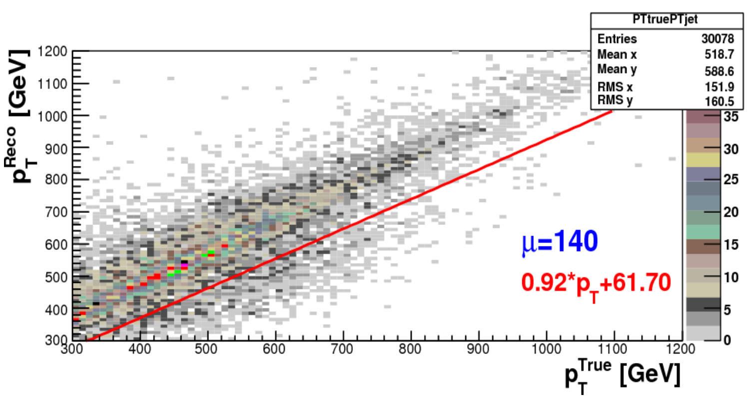
Objects for top reconstruction

- **Electrons, muons and missing ET can be accessed as in the Delphes3 examples**
  - see the "examples/" directory
- **Jets are more difficult to use when using pileup events**
  - Use "Jet" collection
  - Identify "signal" jet by matching it with true jet built using non-pileup particles
    - can mimic jet-vertex fraction algorithm used in ATLAS
  - Apply simple offset jet-energy correction
- **For backgrounds, use parametric jet smearing or samples without truth particle information (small file sizes)**

Resolution studies for high-pileup events







- Generate HERWIG++ signals ($t\bar{t}$) and mix with 50 or 140 pileup events
- Compare antiKT5 jets reconstructed using signal+pileup and antiKT5 jets from truth record (no pileup particles)
- Match truth and reco particles in Eta and Phi (R=0.15)

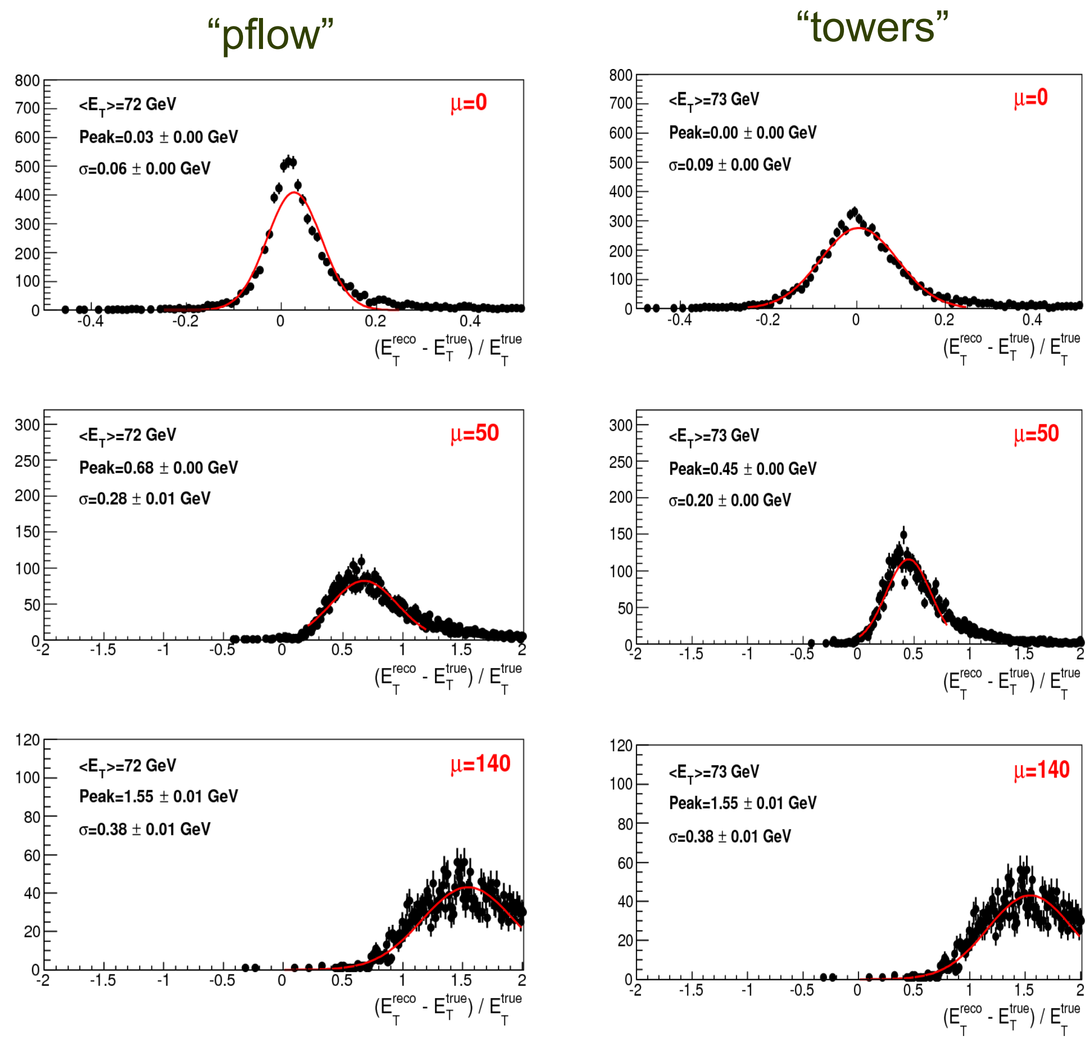
Pileup contribution:

- increase in jet pT (need for an offset correction).
→ shift by ~100 GeV for mu=140
→ *can be reproduced in Delphes without adding pileup MC*
- larger jet smearing (event-by-event correction)
→ *apply a jet smearing without adding pileup MC*
- presence of fake jets
→ *cannot be done in Delphes without adding pileup MC*

The goal of this study is to look at Delphes+realistic pileup events and compare with the full simulation

Jet resolution studies for high-pileup events

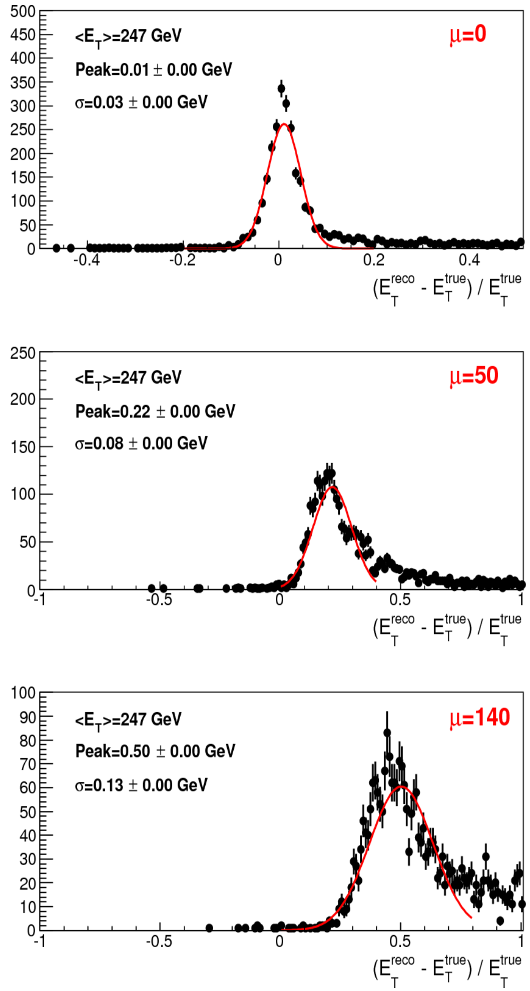
- Default Delphes3 ATLAS geometry card for antiKT4 and "pflow"
  - similar to CMS jets
- Modified Delphes3 ATLAS geometry for antiKT4 using "towers"
  - mimics ATLAS jets



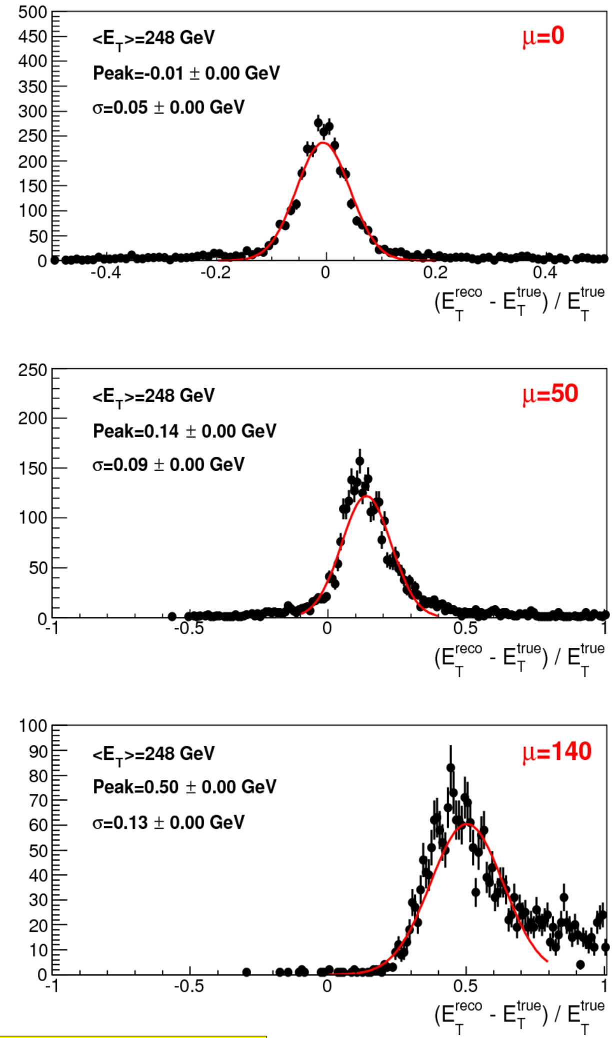
jet resolution for "pflow" looks better for low pT

Resolution studies for high-pileup events

"pflow"



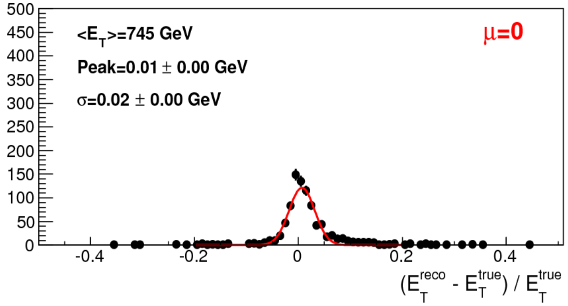
"towers"

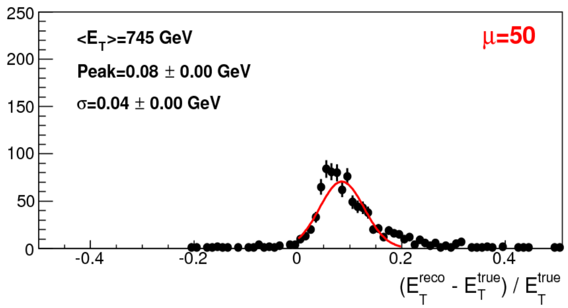


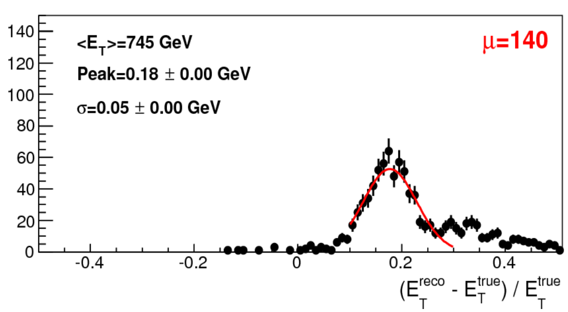
similar performance

Resolution studies for high-pileup events

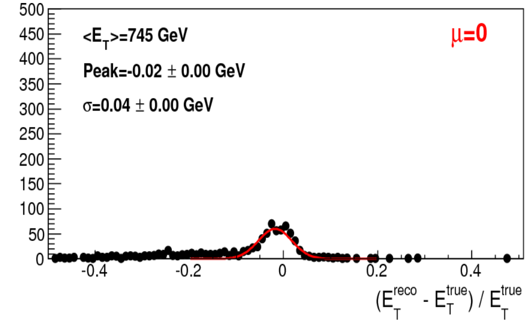
“pflow”

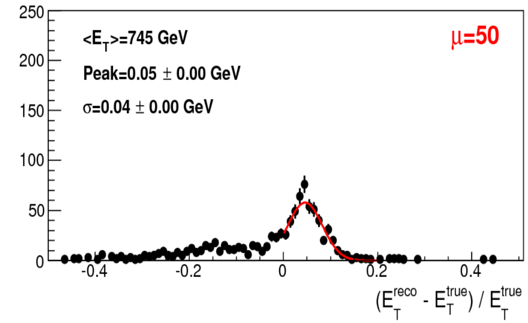


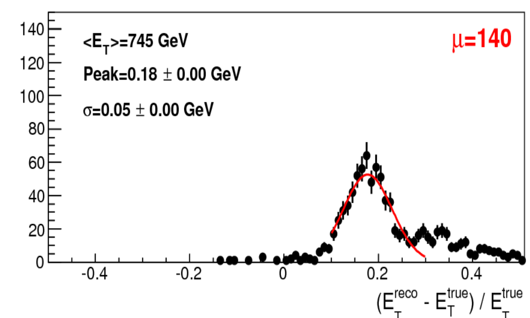




“towers”

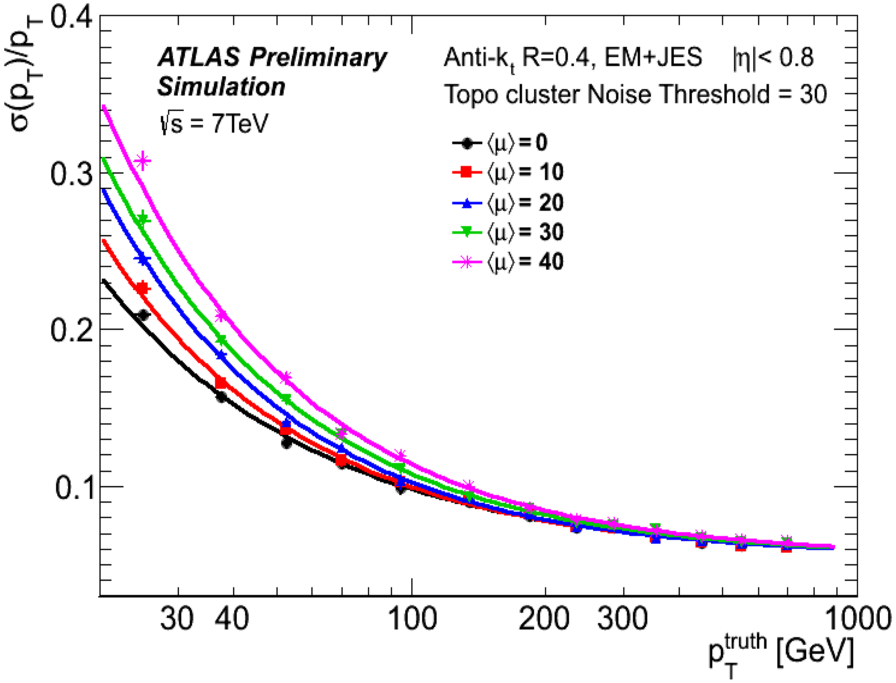






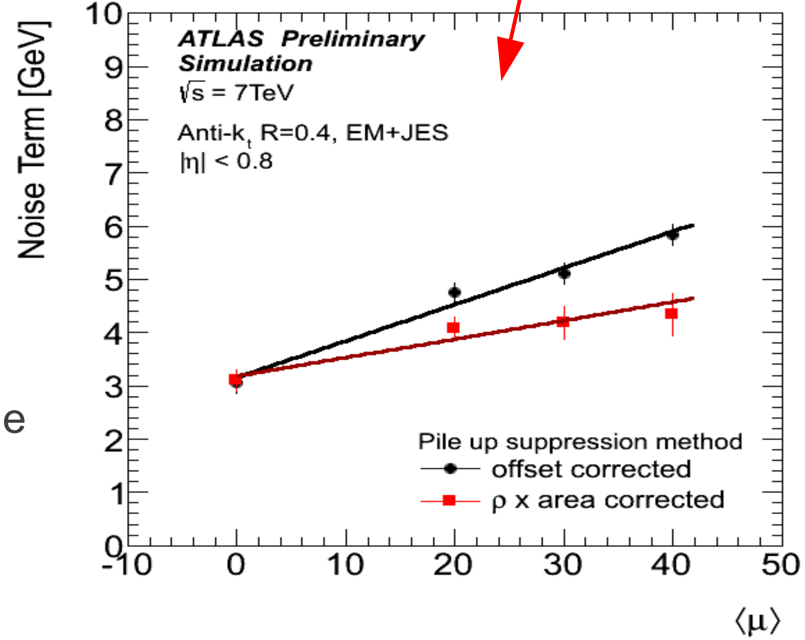
similar performance

ATLAS jet resolution vs <mu> (full simulation)



$$\frac{\sigma(E)}{E} = \sqrt{\frac{a^2}{E} + \frac{b^2}{E^2} + c^2}$$

"Noise term"



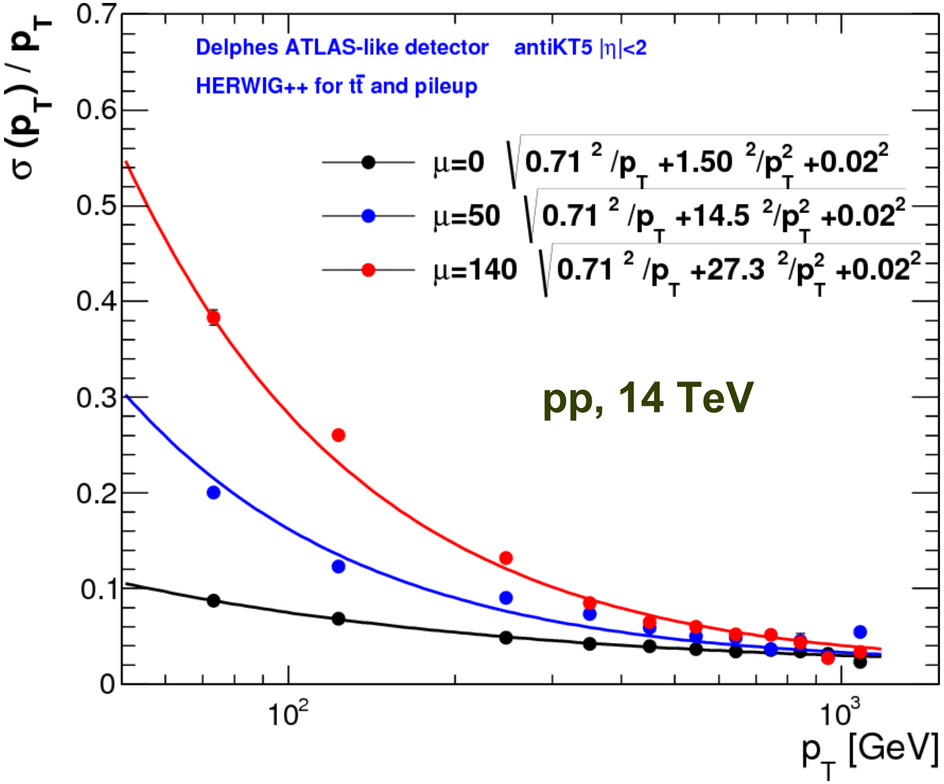
- Keep "sampling" and "constant" terms to be the same when fitting <μ> ≠ 0 cases
- Noise terms increase linearly with <μ>

Extrapolated noise term at <μ> =150:
14 GeV (average offset)
8 GeV (jet area)

Presented by A.Schwartzman at Joint Snowmass-EuCARD/AccNet-HiLumi LHC meeting 'Frontier Capabilities for Hadron Colliders 2013'

Delphes fast simulation. Jet resolution vs <μ>

Delphes tower jets



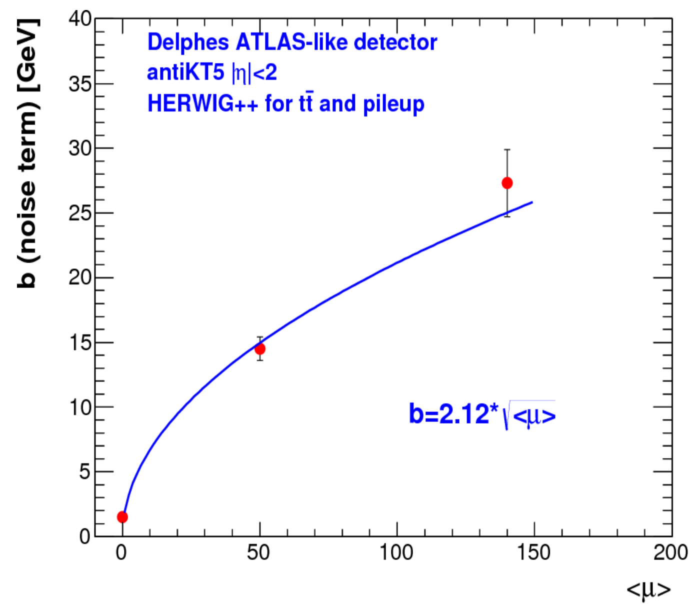
$$\frac{\sigma(E)}{E} = \sqrt{\frac{a^2}{E} + \frac{b^2}{E^2} + c^2}$$

Delphes result agrees with the assumption that pileup mainly changes the noise term ("b")

Noise term ("b") for high-pileup scenario:

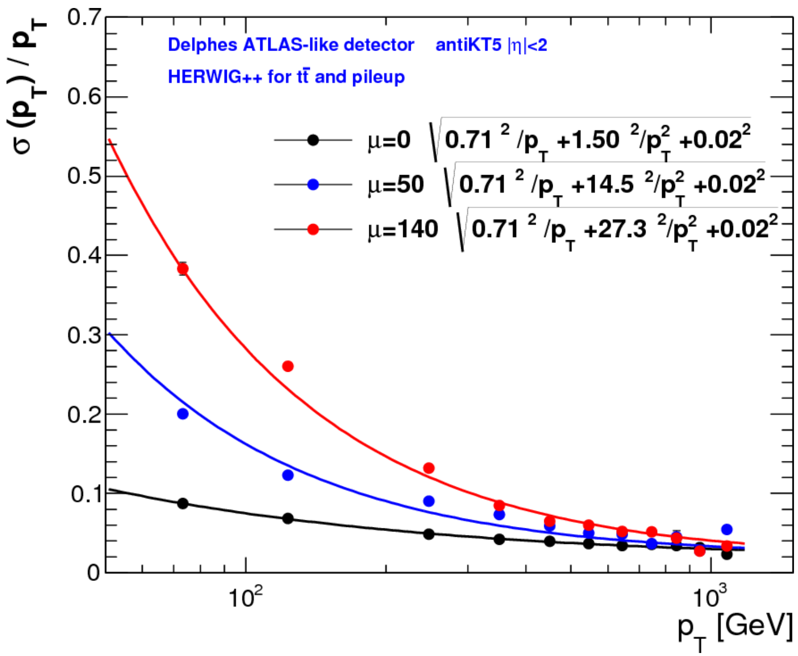
Delphes: b ~ 27 GeV for <μ>~140 for 14 TeV

ATLAS full simulation (extrapolation):
b ~14 GeV for <μ>~150 for 7 TeV

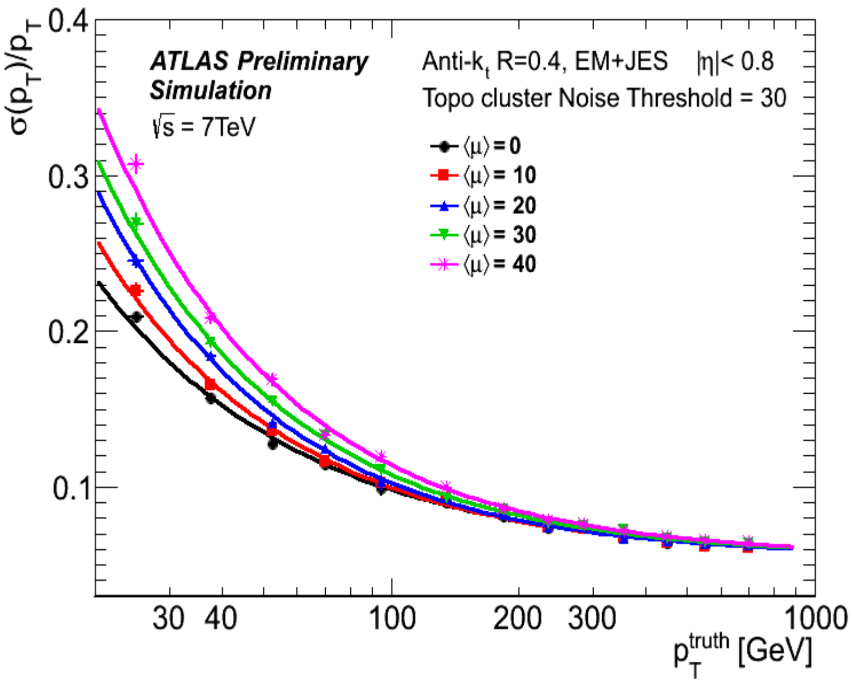


Is Delphes fast simulation realistic for <μ>=0?

Delphes tower jets



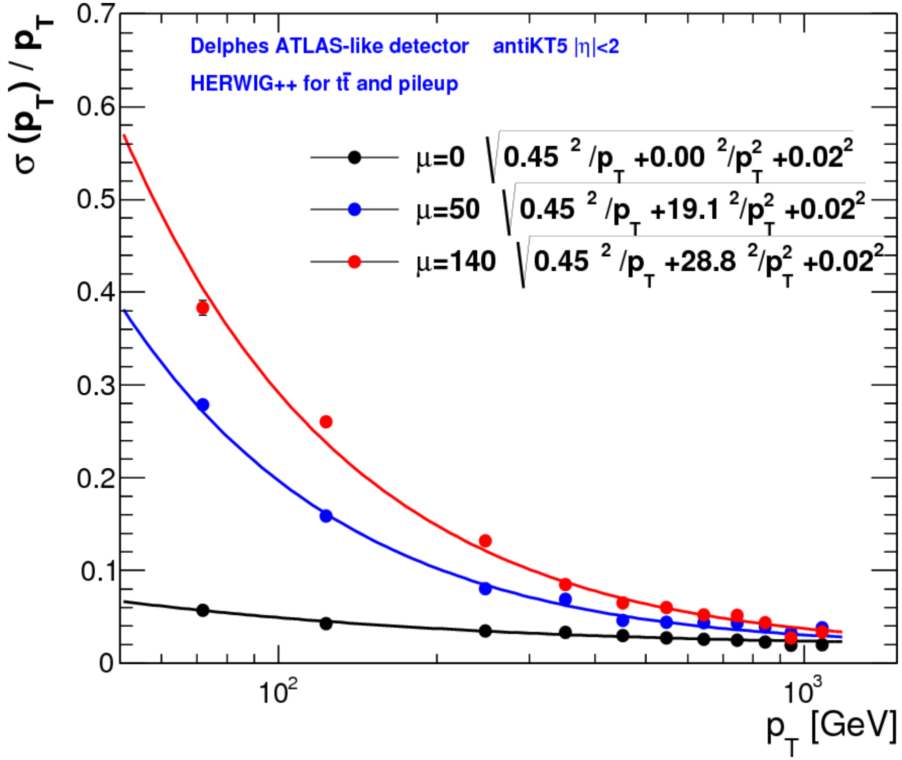
ATLAS Full simulation



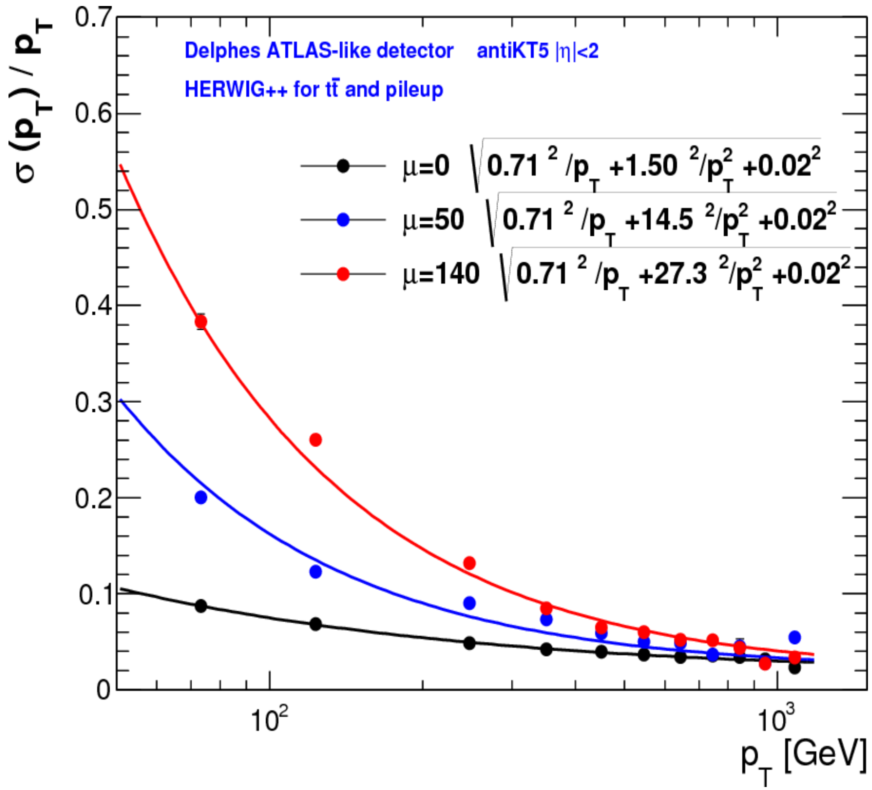
Delphes3 has ~ similar resolution for ATLAS for <mu>=0

Jet resolution studies

"pflow"



"towers"



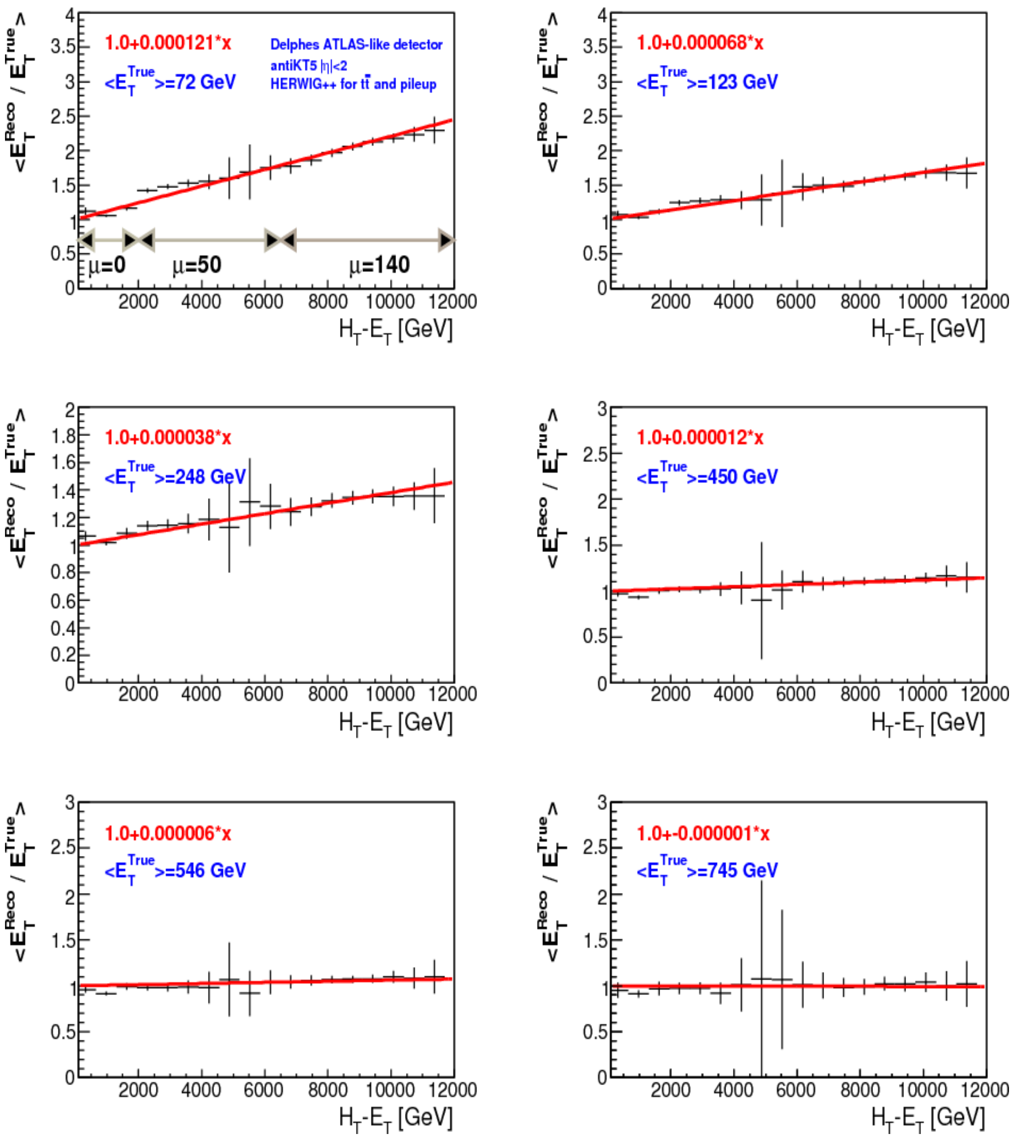
Delphes "pflow" + ATLAS geometry has smaller sampling term than "ATLAS "towers"
The noise term is somewhat larger for the pflow method when μ > 0
Jet resolution for "Towers" and "Pflows" are similar for μ > 0
Performance of pileups subtraction techniques will be essential for proper comparison

Pileup correction

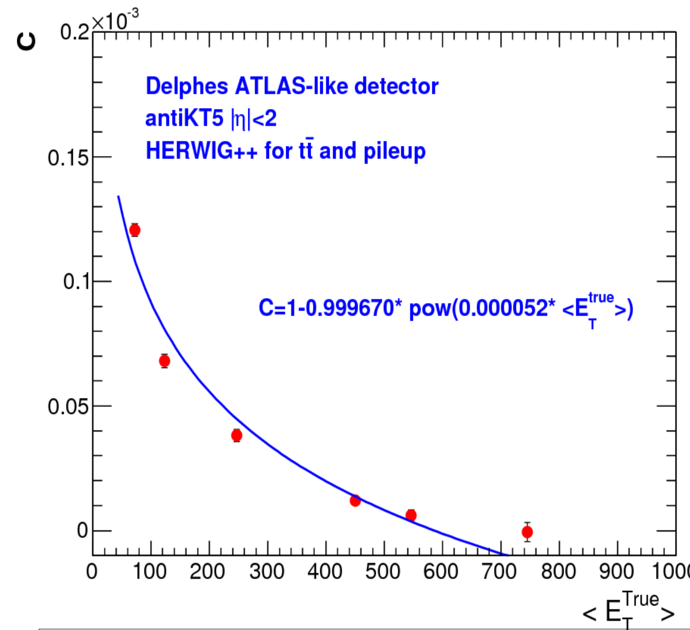
- In ATLAS, jet vertex fraction and jet area corrections are used to subtract pileup
- Jet area correction will be used for the "Snowmass" detector
- Meanwhile, one can use the scalar H_T provided by Delphes3 for a simple offset correction
- **Such correction is provided:**

$$\boxed{E_T^{corr} = E_T / (1 + C \times (H_T - E_T))}$$

Pileup offset correction



- $E_T^{corr} = E_T / (1 + C \times (H_T - E_T))$

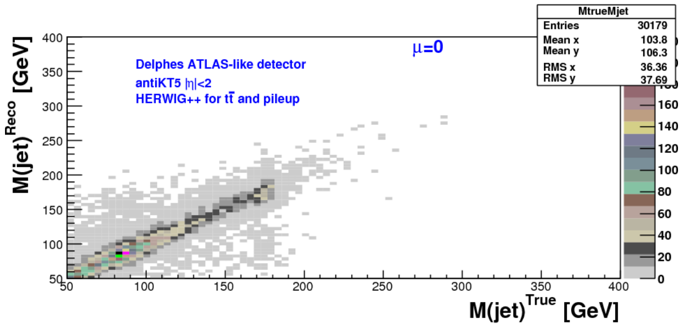


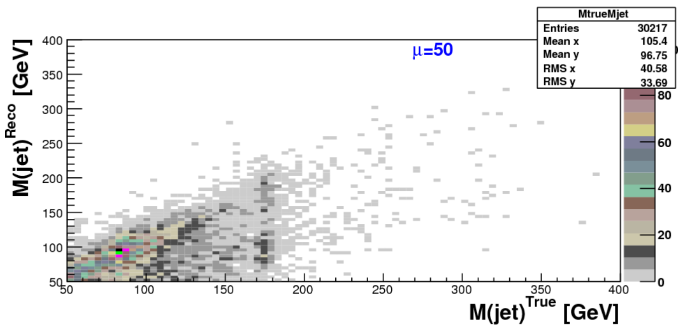
Can be used for physics case studies using Delphes+pileup

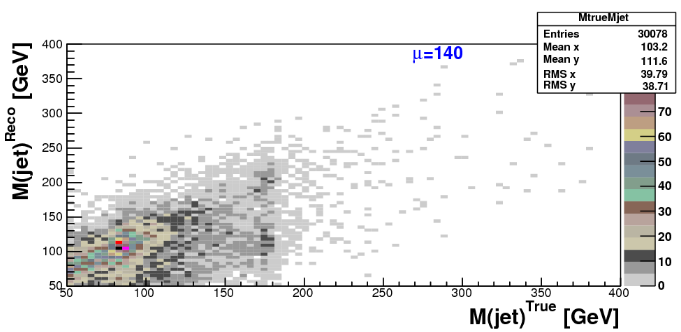
```
double correctJet (Jet *jet, double HT ) {
  double oldPT=jet->PT;
  double C=1.0-0.999670*pow(oldPT,5.18488e-05);
  return oldPT/(1.0+C*(HT-oldPT));
}
```

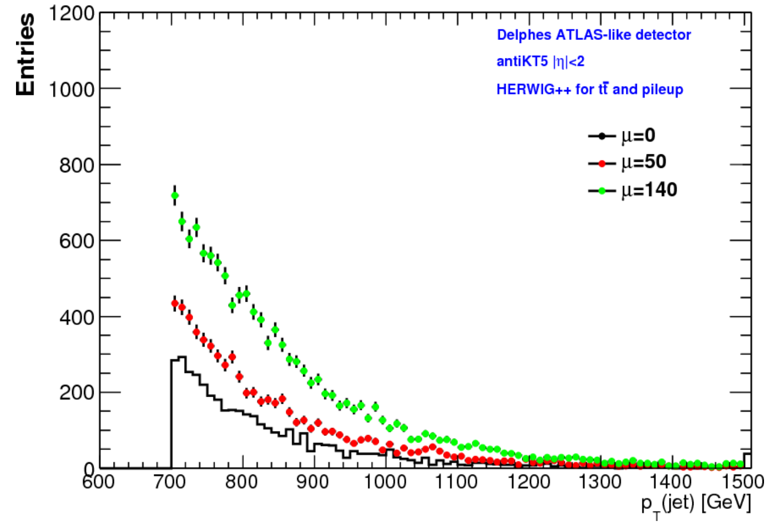
Few physics cases
(no pileup correction applied)

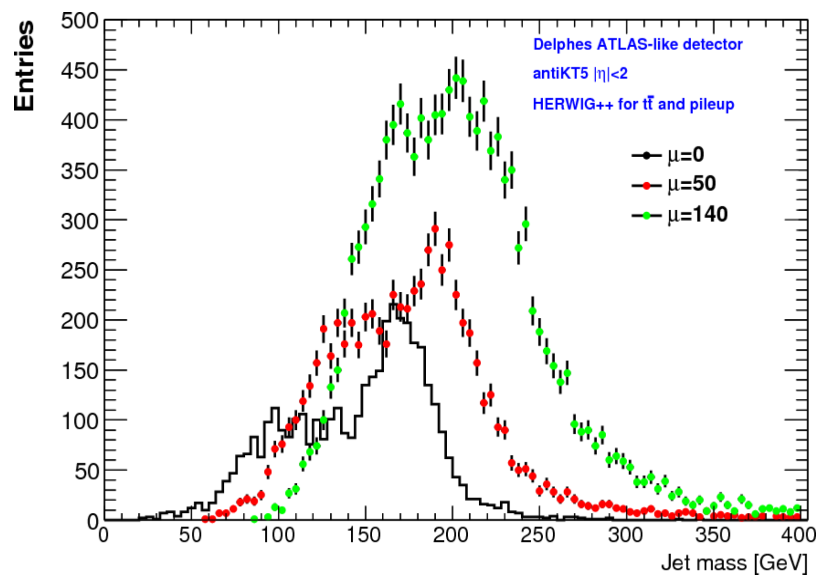
Jet mass (pT>700 GeV) for $t\bar{t}$ events (antiKT5) ("boosted top")



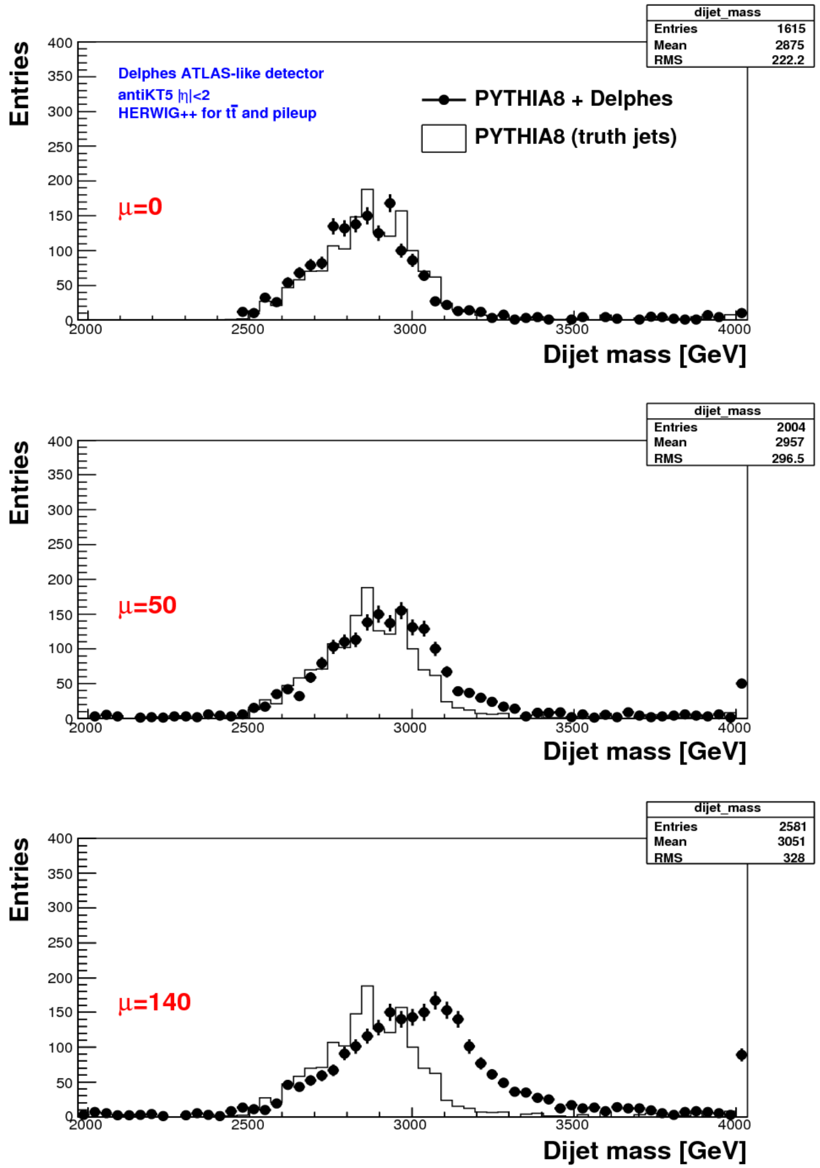








Z' → $t\bar{t}$ with M(Z')=3000 GeV



- **Dijet invariant mass using pT(jet)>1200 GeV and |eta|<2**
- **Mass shift due to pileup events:**
  - mu=50: ~80 GeV
  - mu=140: ~175 GeV